



**QUEEN'S
UNIVERSITY
BELFAST**

“Pattern Recognition” special issue: Sparse representation for event recognition in video surveillance

Zhou, H., Zhang, J., Wang, L., Zhang, Z., & Brown, L. M. (2013). “Pattern Recognition” special issue: Sparse representation for event recognition in video surveillance. *Pattern Recognition*, 46(7), 1748-1749. <https://doi.org/10.1016/j.patcog.2013.01.026>

Published in:
Pattern Recognition

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2015, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/> which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Editorial

“Pattern Recognition” special issue: Sparse representation for event recognition in video surveillance

Understanding and interpreting object behaviours based on video analysis has witnessed tremendous progress in the last decades. However, the performance of state of the art systems still needs to be significantly improved, and many challenges remain to be solved. One of these challenges is that there is a gap between the primitive information directly extracted from images and semantic representations. To bridge this gap, a number of approaches (e.g., bag of words) based on features (e.g. SIFT, HoG and SURF) have been reported to address the coherence between the extracted features and their semantic interpretations. Unfortunately, due to feature redundancy and complexity, these primitive features do not always lead to consistent or semantically meaningful representations.

In recent years, sparse signal representation has proven to be an effective tool for representing and compressing images. It has been demonstrated that optimal decomposition of the signals using appropriately designed bases (e.g. wavelets, Fourier basis or spatio-temporal words) can lead to less ambiguity/uncertainty and computational complexity in image interpretation and representation. In spite of its success, sparsity analysis still has limitations in terms of computational efficiency and interpretation accuracy. As a result, continuous efforts are currently being dedicated towards adaptive sparsity transformations of morphological data such as extracted events in video surveillance.

The primary purpose of this special issue is to organise a collection of recently developed analysis techniques based on sparse representations for understanding/interpreting object behaviours in videos. It provides an international forum for researchers to report and share the recent original developments in this field in an original research paper style. This special issue includes 12 papers which have been accepted in the categorisation of five themes: human detection and tracking, human behaviour analysis, scene analysis, event recognition and face recognition. See details as follows:

- (1) **Human detection and tracking:** This is an active research topic in the last decade. The major challenge in the recent development is how to maintain the performance of the systems in partial occlusions, pose or illumination changes. In this special issue, *Zhang, Li, et al.* introduced a new L1-regularised sparse representation based multi-target tracking system. This system used block-division based covariance features rather than image patches for better discrimination. In addition, a subtle template dictionary, consisting of a fixed template, a stable template and variational templates, was constructed to capture the variations of the appearance. *Lu et al.* proposed a non-local self-similarity (NLSS) based sparse coding algorithm for object tracking. This algorithm considered the geometrical structures of the tracking candidates and modelled their changes over time.

Zhang, Yao, et al. reviewed the state of the art tracking methods based on sparse coding. The available systems can be categorised into appearance modelling based on sparse coding (AMSC) and target searching based on sparse representation (TSSR). The comparison conducted shows that the AMSC methods outperform the TSSR methods in terms of tracking accuracy.

- (2) **Human behaviour analysis:** Due to the high dimensionality of human motion, it becomes extremely difficult to recognise human behaviours in a dynamic environment. In this special issue, some authors have made significant efforts to handle this problem. For example, *Zhang, Yao and Wang* presented a model for learning atomic action based on the co-occurrence and temporal proximity in the same activity category, using an extension of Hierarchical Dirichlet Process (HDP) mixture model.

Lu and Peng developed a graph-based spectral embedding approach for latent semantic learning, where the graph was constructed using sparse representation to capture the relationship of mid-level features for action recognition. *Liu et al.* reported a method for human action recognition, based on boosted key-frame selection and correlated pyramid motion feature representations. *Zhang, Yang, et al.* presented a sparse coding scheme that can discover the manifold structure of the learned features with an L2-norm regularisation. The action recognition problem was casted as a sparse linear representation of an over-complete dictionary constricted by labelled behaviour representations. *Wang et al.* presented an instant action recognition method, using optical flow and edges, in combination with a boosting classifier.

- (3) **Scene analysis:** scene understanding and analysis has drawn the major attention of the computer vision community in recent years. This is mainly driven by the applications in video surveillance. *Fu et al.* introduced an unsupervised approach to explore motion patterns that occur in the scenes using an improved sparse topical coding (STC) framework. An input video was firstly segmented into a sequence of clips without overlapping. Afterwards, visual flow words were generated using optical flows. The STC was then applied to extract latent patterns that represent the common motion distributions of the scene.

Liu et al. investigated the geometrical structures of the features in two situations. In a simple case, the features were classified using locally linear embedding with k-nearest neighbours. In a real-world scenario, a sparse representation algorithm was used, combined with k-nearest neighbours. These two structures were then embedded into sparse coding algorithm, which was solved by a coordinate descent scheme.

- (4) **Event recognition:** In this special issue, *Cong et al.* proposed to detect abnormal events by a sparse reconstruction over the normal basis. A sparse reconstruction cost (SRC) was introduced to measure the normality of the testing samples. A dictionary selection method with group sparsity constraint was designed to condense the over-completed normal basis into a compact dictionary.
- (5) **Face recognition:** one of the major issues in the research development of face recognition is systematic performance degradation in the presence of occlusion. In our special issue, a good attempt has been made to deal with this problem. *Yang et al.* proposed a Gabor feature based representation and classification scheme for face recognition. Using the Gabor features, L2-norm can be used to regularise the coding coefficients which reduced the computational cost in coding occluded face images.

Acknowledgement

The Guest Editors would like to thank all the authors for submitting their high quality manuscripts to this special issue, and all the reviewers for providing their timely review comments and suggestions. They would also like to thank the Editor-in-Chief Prof. Ching Y.

Suen and the journal managers Ms. Carol Bergin and Sally Woods for encouraging and supporting this special issue.

Guest editors of this special issue:

Huiyu Zhou, PhD, Institute of Electronics, Communications and Information Technology, Queen's University Belfast, United Kingdom. E-mail: h.zhou@ecit.qub.ac.uk.

Jianguo Zhang, PhD, School of Computing, University of Dundee, United Kingdom. E-mail: jgzhang@computing.dundee.ac.uk.

Liang Wang, PhD, National Lab of Pattern Recognition, Beijing, China. E-mail: wangliang@nlpr.ia.ac.cn.

Zhengyou Zhang, PhD, Microsoft Research, Microsoft Corp., One Microsoft Way, Redmond WA, USA. E-mail: zhang@microsoft.com.

Lisa M. Brown, PhD, IBM Thomas J. Watson Research Center, Hawthorne, NY USA. E-mail: lisabr@us.ibm.com.